

Factors Governing the Expression of a Bacterial Gene in Mammalian Cells

RICHARD C. MULLIGAN† AND PAUL BERG*

Department of Biochemistry, Stanford University Medical Center, Stanford, California 94305

Received 12 February 1981/Accepted 10 March 1981

Cultured monkey kidney cells transfected with simian virus 40 (SV40)-pBR322-derived deoxyribonucleic acid (DNA) vectors containing the *Escherichia coli* gene (Ecogpt, or *gpt*) coding for the enzyme xanthine-guanine phosphoribosyltransferase (XGPRT) synthesize the bacterial enzyme. This paper describes the structure of the messenger ribonucleic acids (mRNA's) formed during the expression of *gpt* and an unexpected feature of the nucleotide sequence in the *gpt* DNA segment. Analyses of the *gpt*-specific mRNA's produced during infection of CV1 cells indicate that in addition to the mRNA's expected on the basis of known simian virus 40 RNA splicing patterns, there is a novel SV40-*gpt* hybrid mRNA. The novel mRNA contains an SV40 leader segment spliced to RNA sequences transcribed from the bacterial DNA segment. The sequence of the 5'-proximal 345 nucleotides of the *gpt* DNA segment indicates that the only open translation phase begins with an AUG about 200 nucleotides from the end of the *gpt* DNA. Two additional AUGs as well as translation terminator codons in all three phases precede the XGPRT initiator codon. Deletion of the two that are upstream of the putative start codon increases the level of XGPRT production in transfected cells; deletion of sequences that contain the proposed XGPRT initiator AUG abolishes enzyme production. Based on the location of the XGPRT coding sequence in the recombinants and the structure of the mRNA's, we infer that the bacterial enzyme can be translated from an initiator AUG that is 400 to 800 nucleotides from the 5' terminus of the mRNA and preceded by two to six AUG triplets.

A principal goal of our laboratory has been to adapt simian virus 40 (SV40) as a vector for introducing genes into mammalian cells (7, 8, 21-23). One effort was directed at developing vectors which promote the expression of complementary deoxyribonucleic acid (cDNA) or other protein coding sequences, since these lack sufficient information to effect their own transcription and processing. After recognizing that splicing was needed to ensure maximal accumulation of SV40 late messenger ribonucleic acids (mRNA's) (8, 10), it became possible to engineer the expression of a variety of sequences, including cDNA's for rabbit β -globin (22), rabbit α -globin (H. Okayama and P. Berg, manuscript in preparation), mouse dihydrofolate reductase (S. Subramani, R. C. Mulligan, and P. Berg, manuscript in preparation), and bovine parathyroid hormone (R. C. Mulligan, H. M. Kronenberg, A. Rapoport, I. Tepler, J. T. Potts, A. Rich, and P. Berg, unpublished data). Success in these

cases depended upon specific SV40 DNA sequences for the initiation, termination, and maturation of SV40 RNA transcripts and some inferences about the optimal placement of protein coding sequences within an mRNA for efficient translation.

The availability of cloned procaryote DNA segments coding for unique products provides a potentially rich source of genes for introduction into mammalian cells. Accordingly, we examined the feasibility of transducing animal cells with bacterial genes. For a variety of reasons, the *Escherichia coli* gene (Ecogpt, or *gpt*) encoding xanthine-guanine phosphoribosyltransferase (XGPRT) (12) seemed a promising candidate. *E. coli* XGPRT resembles the animal cell hypoxanthine-guanine phosphoribosyltransferase (HGPRT) in that it utilizes hypoxanthine and guanine in the condensation with phosphoribosyl pyrophosphate to form inosinic and guanylic acids, respectively (16, 20). However, in contrast to the mammalian HGPRT (16) bacterial XGPRT also catalyzes an efficient conversion of xanthine to xanthine 5'-monophosphate (20).

The isolation of the *gpt* gene by using standard

† Present address: Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, and Endocrine Unit, Massachusetts General Hospital, Boston, MA 02114.

recombinant and genetic techniques, its recombination with several SV40 vectors, and its expression in a variety of cultured mammalian cells have already been described (21). In this paper, we examine the structure of the *gpt*-specific mRNA's formed after transfection of monkey cells with several vector-*gpt* recombinant DNAs to help clarify the mode of their expression. The results suggest that initiation sites for translation of bacterial genes can be recognized in mammalian cells, even when these signals are distant from the 5' terminus of the RNA and preceded by other AUGs.

MATERIALS AND METHODS

Cell culture and DNA transfection. CV1 cells were maintained in Dulbecco-modified Eagle medium supplemented with 5% fetal bovine serum (Microbiological Associates), penicillin, and streptomycin. Twenty-four hours after seeding 10^6 cells on 100-mm plates, the cultures were transfected with 10 μ g of the appropriate pSV-*gpt* DNA by using the procedure of Graham and Van der Eb (9) as modified by Parker and Stark (24).

Electrophoretic assay of purine phosphoribosyltransferase. Seventy-two hours after transfection of CV1 cells with pSV-*gpt* DNAs, cell extracts were prepared as described earlier (21) and electrophoresed on 7.5% acrylamide gels buffered with 0.2 M tris(hydroxymethyl)aminomethane (pH 8.5). The gels were assayed in situ for GPRT activity by a modification (21) of the method of Chasin and Urlaub (4).

Analysis of pSV-*gpt* RNAs. The isolation of cytoplasmic polyadenylate-containing RNAs from CV1 cells and their analysis (1) by electrophoresis in denaturing gels, transfer to nitrocellulose sheets, and hybridization with appropriate labeled probes have been described previously (22). The spliced structure of the RNAs was analyzed by Weaver and Weissmann's modification (28) of the procedure of Berk and Sharp (2); end-labeled hybridization probes (indicated in the text) were prepared with restriction fragments from which the 5'-terminal phosphate residues had been removed with calf intestinal phosphatase (Boehringer-Mannheim) and then restored with adenosine 5'-[γ - 32 P]triphosphate (New England Nuclear Corp.) and T4 polynucleotide kinase (Boehringer-Mannheim) (18). After hybridization of the RNA with an excess of the appropriate labeled DNA fragment, the RNA:DNA hybrid was digested with S1 nuclease (Boehringer-Mannheim), precipitated with ethanol, denatured in NaOH, and electrophoresed in a 1.5% alkaline agarose gel (30 mM NaOH, 1 mM ethylenediaminetetraacetic acid).

Nucleotide sequence of *gpt* DNA. pSV2-*gpt* DNA (see Fig. 1) was cleaved with *Hind*III or *Bgl*II endonuclease (Bethesda Research Laboratories) and then treated with calf alkaline phosphatase, and the 5' ends were labeled with T4 polynucleotide kinase and adenosine 5'-[γ - 32 P]triphosphate. The resulting 32 P-labeled linear DNAs were digested with *Bam*HI endonuclease, and the *Hind*III end-labeled *gpt*-containing fragment (1 kilobase [kb]) and both end-labeled *Bgl*II

fragments (0.9 and 4.2 kb) were analyzed by the Maxam-Gilbert nucleotide sequencing technique (18).

***gpt* complementation assay.** *E. coli* GP120 (*gpt purE*) (12), obtained from J. Gots, University of Pennsylvania, Philadelphia, was transformed by various plasmid DNAs in the standard way (17). Ampicillin-resistant transformants were selected on L-broth plates containing 50 μ g of ampicillin per ml; then, after purification, each colony was tested for the ability to grow on a medium containing minimal salts and guanine. The *gpt hpt*⁺ *E. coli* strains (such as GP120) grow slowly on this medium, but the Gpt⁺ phenotype is clearly recognizable by substantial colony growth in 24 to 36 h (12).

Construction of *gpt* deletions. pSV2-*gpt* DNA was digested with *Bgl*II or *Kpn*I endonuclease to completion and incubated with *E. coli* *Poll* (a gift of S. Scherer, Stanford University, Stanford, Calif.) and the four deoxynucleoside triphosphates for 45 min at 15°C as already described (22). *Hind*III linkers (Collaborative Research), in 40-fold molar excess, were ligated to the blunt-ended termini via ligation with T4 DNA ligase (gift of S. Scherer) (11). After inactivation of the T4 DNA ligase and digestion with *Hind*III endonuclease, the reaction mixture was diluted to a DNA concentration of 2 μ g/ml and incubated with T4 DNA ligase for 24 h at 15°C. Under these conditions, intramolecular joining of the ends to form circular DNAs is the predominant reaction. *E. coli* HB101 was transfected with various dilutions of the ligated DNA, and transformants were obtained on plates containing ampicillin. The plasmid DNA obtained from a number of transformants (5) was screened for the desired structures by cleavage with appropriate restriction enzymes and electrophoresis in 1.2% agarose gels.

RESULTS

Transfecting vectors for *gpt* expression. The construction and propagation of SV40 recombinant genomes containing a rabbit β -globin cDNA sequence in place of various portions of the late region has already been reported (22, 23). Analogous recombinants have also been constructed with a mouse dihydrofolate reductase cDNA (3) as the inserted segment (Subramani et al., in preparation). Infection of cells with these recombinant genomes yields RNAs whose structures are consistent with those predicted on the basis of SV40 mRNA structures, i.e. RNAs corresponding in size and structure to SV40 16S and 19S late mRNA's; moreover, the polypeptides encoded by the cDNA's are also synthesized.

In applying this strategy to the cloning of the *gpt* DNA (21), we found it difficult to propagate the virions containing hybrid SV40-*gpt* genomes; generally, plaque-purified stocks of SVGT5-*gpt* contained predominantly the helper virus (*tsA58*), variable but low amounts of the expected recombinant genome, and recombinant genomes with aberrant structures (Mulligan and

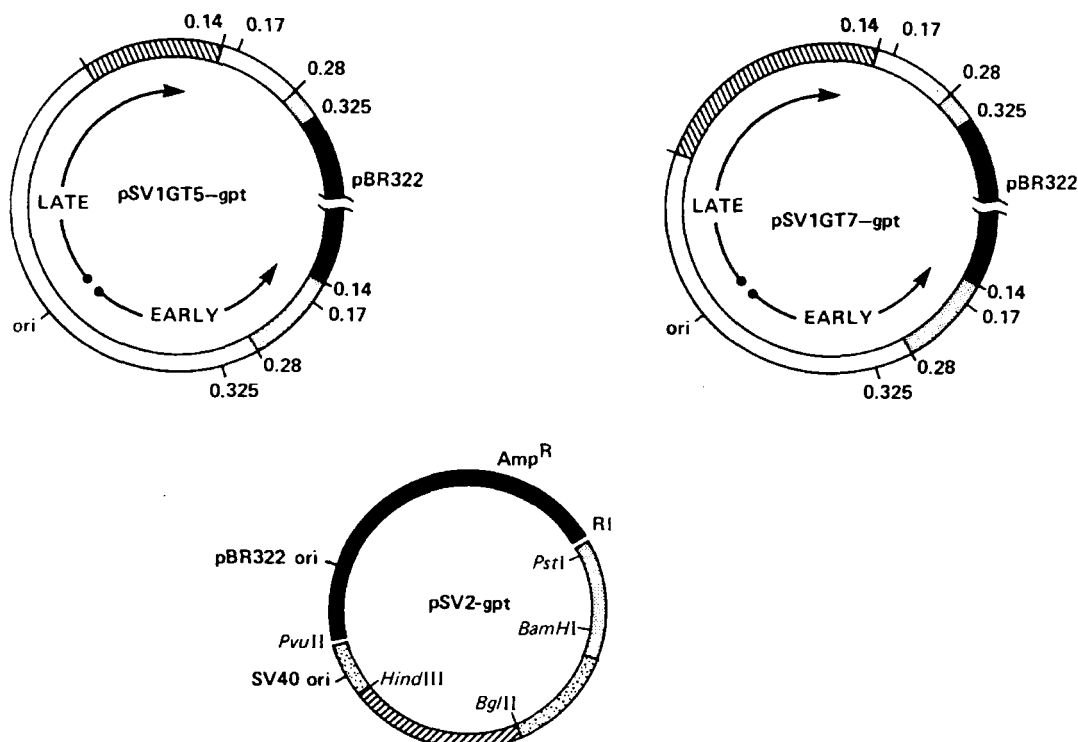


FIG. 1. Structure of pSV-gpt recombinants. All the vectors are shown with the *gpt* segment hatched. The solid black segments in each of the diagrams represent pBR322 DNA sequences. In pSV1GT5- and pSV1GT7-gpt, the pBR322 segment is the 4-kb fragment that extends from its unique *Hind*III restriction site to the unique *Bam*HI site, whereas in pSV2-gpt, it is the 2.3-kb fragment that extends from the unique *Eco*RI restriction site to the unique *Pvu*II restriction site. SV40 sequences in pSV1GT5- and pSV1GT7-gpt are shown as open or shaded portions of the circle, and those in pSV2-gpt are shown as stippled or shaded regions.

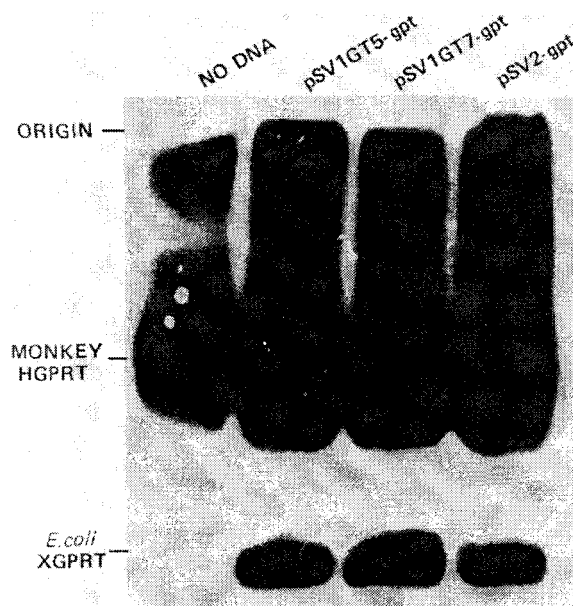


FIG. 2. Assay of extracts from pSV-gpt-transfected cells for GPRT activity after electrophoresis in polyacrylamide gels. The positions of CV1 HGPRT and *E. coli* XGPRT are indicated. Each extract sample corresponds to approximately 10^6 cells.

Berg, unpublished data). The reasons for this difficulty are not clear, but virus production may be reduced in cells that are amplifying recombinants with an expressible *gpt* gene. This problem was averted by the use of recombinants that contain the desired SV40-*gpt* structures inserted into pSV1, an SV40-pBR322 hybrid DNA which can be propagated and maintained as a plasmid in *E. coli*. (The details of the construction and characterization of pSV1 and its derivatives are being published elsewhere.) Each of the resulting recombinant plasmids, pSV1GT5-*gpt* and pSV1GT7-*gpt* (Fig. 1), retains an intact early region, which permits them to replicate in animal cells, as well as those portions of the late region needed for the synthesis of late RNAs. In pSV1GT5-*gpt*, the *gpt* segment replaces the coding sequence for the major capsid protein, VP1; in pSV1GT7-*gpt*, some of the coding region for the minor capsid proteins VP2 and VP3 is deleted as well. The *gpt* gene has also been inserted into another vector, pSV2 (Fig. 1) (B. H. Howard, P. J. Southern, and P. Berg, manuscript in preparation) so that the formation of XGPRT is mediated by SV40 early rather than late transcription. Because pSV1GT5-*gpt* and pSV1GT7-

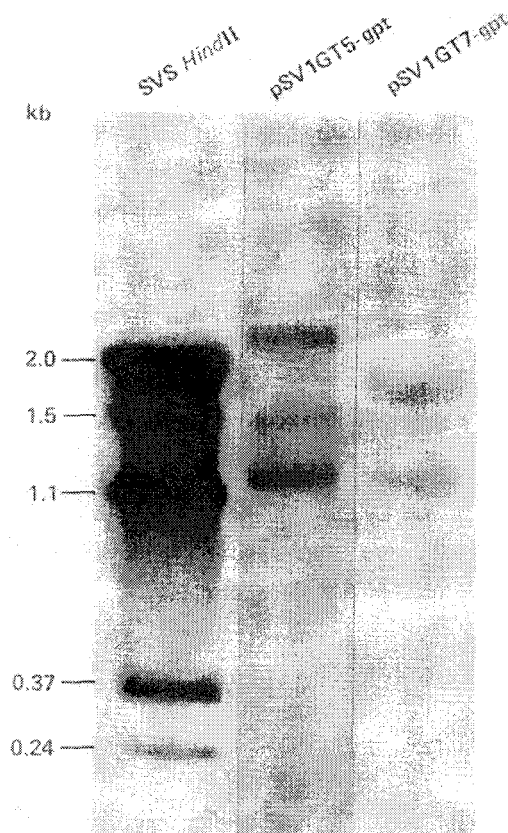


FIG. 3. Analysis of pSV-*gpt* RNAs. Sizes of *gpt*-containing mRNA's. See the text for reference to the electrophoretic, blotting, and hybridization procedures. The probe for the blot hybridization was ^{32}P -labeled pBR322-*gpt* DNA. *Hind*II-cleaved SV40 DNA provides the size markers. Each track contained RNA extracted from approximately 4×10^7 transfected cells.

gpt contain more than 9 kb of DNA, they cannot be incorporated into SV40 virions. Nevertheless, these plasmids can be propagated in *E. coli*, and the DNA obtained from such bacterial cultures can be introduced into CV1 cells by using standard techniques of DNA transfection (9, 24) (see Materials and Methods). Although only about 5 to 10% of the cells are transfected with the plasmid DNAs under these conditions, expression of the bacterial enzyme can be detected by the appearance in the cell extracts of GPRT activity with an electrophoretic mobility characteristic of *E. coli* XGPRT (Fig. 2). Such transfected cells have been used to examine the structure of the *gpt*-specific mRNA's responsible for XGPRT production.

Production of SV40-*gpt* RNA in pSV-*gpt*-transfected cells. Our present understanding of SV40 transcription permits certain predictions about the structure of the mRNA species

that would be produced by the above recombinants. pSV1GT5-*gpt* retains all of the noncoding nucleotide sequences involved in the biogenesis of SV40 late RNAs: the SV40 late promoter and the sites involved in RNA splicing, transcription termination, and polyadenylation. Only the region encoding the SV40 major capsid protein, VP1, has been replaced by the *gpt* DNA segment. It would be expected, therefore, that the expression of pSV1GT5-*gpt* should yield two polyadenylated, *gpt*-containing RNAs corresponding in structure to the 16S and 19S SV40 late mRNA's. Since the 1-kb *gpt* segment is the same size as the excised VP1 coding sequence, the SV40-*gpt* hybrid RNAs should be the same size as the SV40 16S and 19S mRNA's (1.5 kb and 2.3 kb, respectively). With pSV1GT7-*gpt*, however, the site to which the SV40 16S leader (map position 0.72 to 0.76) is spliced (map position 0.939) (6) is absent. Consequently, only one mRNA species (about 1.9 kb), corresponding to the SV40 19S class of late mRNA, should be produced by this hybrid.

These predictions were tested by transfecting CV1 cells with pSV1GT5-*gpt* or pSV1GT7-*gpt* DNA and examining the cytoplasmic polyadenylated RNAs by electrophoresis in 1.5% denaturing agarose gels (19), transfer to diazotized benzyloxymethyl paper (1), and hybridization of the blots with ^{32}P -labeled, nick-translated pBR322-*gpt* DNA (26) (Fig. 3). With pSV1GT5-*gpt*, *gpt*-containing mRNA's of the expected size (1.5 and 2.3 kb) are readily discernible; similarly, transfections with pSV1GT7-*gpt* DNA yield the expected 1.8-kb mRNA. However, in contrast to the result with SV40 DNA in which there is considerably more 16S than 19S mRNA (29), transfections with pSV1GT5-*gpt* produce less 16S- than 19S-like mRNA. Moreover, an unexpected RNA, 1.2 kb in length, is produced in cells transfected with both pSV1GT5-*gpt* and pSV1GT7-*gpt*. This novel RNA contains SV40 late leader sequences, since it also hybridizes with a DNA probe containing sequences from that region (0.67 to 0.76 map unit) (data not shown).

Another facet of the structures of the *gpt*-containing RNAs produced by these recombinants was revealed by using the Weaver-Weissmann modification (28) of the method of Berk and Sharp (2). In this procedure ^{32}P -end-labeled probes, obtained from pSV1GT5-*gpt* and pSV1GT7-*gpt* DNAs (see Fig. 4A), were annealed with RNA from the corresponding transfections; then, the RNA:DNA hybrids were digested with S1 nuclease, and the lengths of the protected ^{32}P -labeled fragments were determined by electrophoresis in alkaline agarose gels

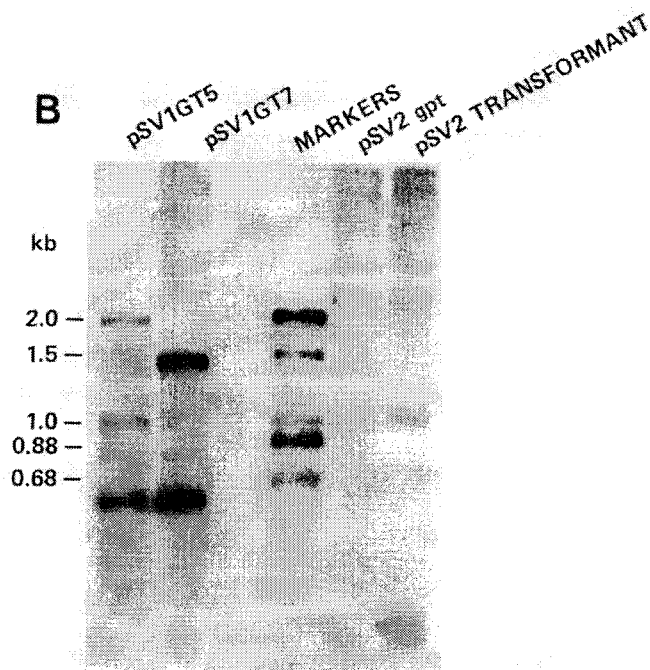
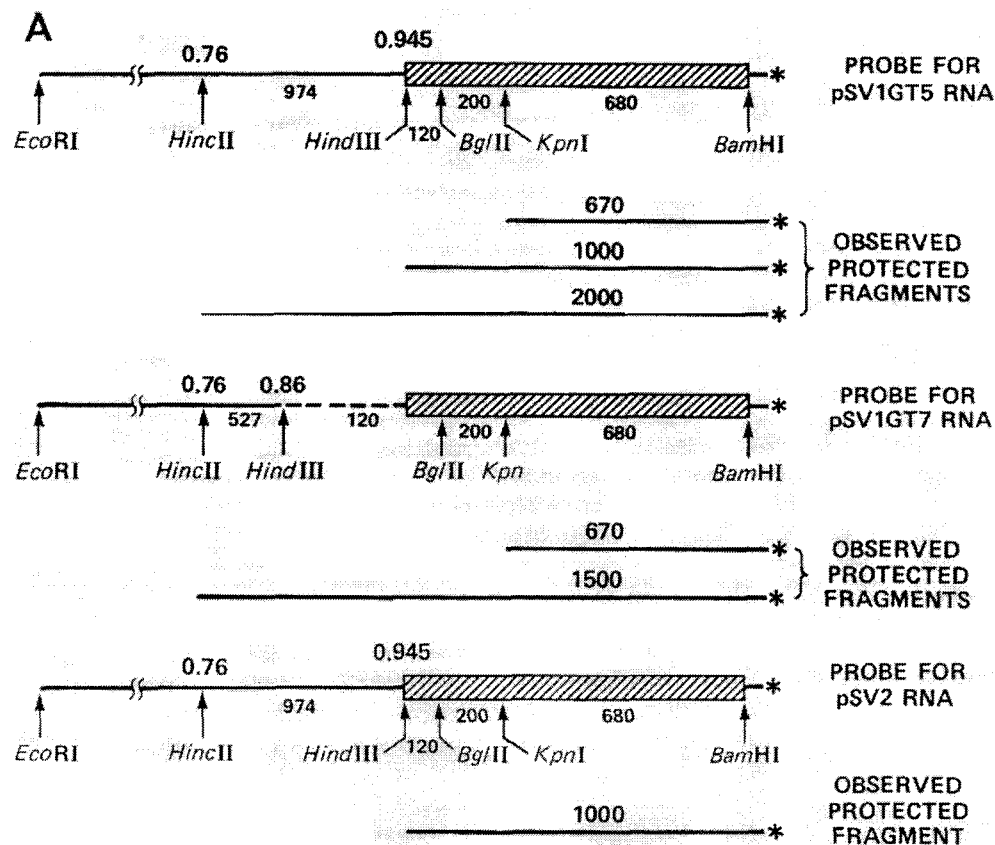


FIG. 4. (A) Splicing of *gpt*-containing mRNA's. The specific end-labeled probes hybridized to the respective RNA samples are indicated with solid lines for SV40 sequences and by the hatched box for *gpt* sequences. The distances in nucleotides between restriction endonuclease sites for a number of enzymes are indicated, as are the lengths of the observed protected fragments. The asterisks indicate the label at the 5' end. (B) After hybridization of the labeled DNA probes to the RNA, the samples were digested with S1 nuclease and electrophoresed (see text for details). The markers, whose sizes are shown in kb, were obtained by cleavage of the 32 P-end-labeled pSV1GT5-*gpt* and pSV1GT7-*gpt* probes with HincII, HindIII, BglII, or KpnI endonucleases. The pSV1GT5-*gpt* and pSV1GT7-*gpt* RNA samples were extracted from 2×10^7 cells; the pSV2-transfected and pSV2-transformed cell RNAs were obtained from about 10^8 cells.

(Fig. 4B). The length of the labeled DNA fragments measures the segment length of RNA that is homologous to the labeled end and a contiguous region of the vector DNA, thereby indicating the site(s) at which splicing has occurred. For pSV1GT5-*gpt* RNA, protected fragments of 1 and 2 kb are consistent with the structures of the expected 16S and 19S-like RNAs. Similarly, the 1.5-kb labeled DNA protected by pSV1GT7-*gpt* RNA is consistent with the expected structure of the 19S-like RNA. The presence of a 0.67-kb protected fragment with RNA from pSV1GT5-*gpt* and pSV1GT7-*gpt* (Fig. 4B) indicates that the 1.2-kb RNA is also spliced, with an SV40 leader segment joined to a sequence within the bacterial DNA segment.

To assess the possibility that an S1 nuclease-sensitive site within the pSV1GT5-*gpt* or pSV1GT7-*gpt* RNA:DNA hybrids accounts for the 0.67-kb protected fragment, RNA obtained from cells transfected with pSV2-*gpt* DNA or from TC7 cells that contain a single integrated copy of pSV2-*gpt* DNA (Mulligan and Berg, unpublished data) was analyzed with the end-labeled probe used for the pSV1GT5-*gpt* RNA. With RNA from either pSV2-*gpt*-transfected or -transformed cells, only a single fragment, 1 kb in length, was protected (see Fig. 4B); this indicates that the entire *gpt* sequence is represented in the RNA produced by pSV2-*gpt* and that no anomalous cleavage by S1 nuclease occurs within that sequence. Our findings and interpretations concerning the structure of the mRNA's produced by pSV1GT5- and pSV1GT7-*gpt* DNAs are summarized in Fig. 5.

5' Nucleotide sequence of *gpt*. At the time *gpt* was cloned, the position of the XGPRT coding sequence within the 1-kb DNA segment was not known. To aid in the identification of which of the *gpt*-containing mRNA's are translated to yield XGPRT, we sought to locate the enzyme's amino-terminal coding sequence. Since the amino acid sequence of the protein is not known, the location of the start of the coding region in the sequence cannot be identified simply. Nevertheless, the sequence of the 5'-proximal 345 nucleotides in the *gpt* DNA segment (Fig. 6A) reveals an open translation phase that is appropriate to code for XGPRT.

The expected nucleotide sequences for the already established *Hind*III, *Bgl*II, and *Kpn*I restriction endonuclease sites are located at nucleotides 1, 121, and 322, respectively. Note that there is a near-perfect Pribnow box (25) straddling nucleotide 127, and promoter-associated sequences about 35 nucleotides upstream from the Pribnow box (27). Perhaps this locates the *gpt* promoter that permits efficient expression of the *gpt* segment from plasmid-*gpt* recombinants propagated in *E. coli* (21). The disposition of AUG and termination codons reveals that there is only one translation phase that could encode XGPRT, that is, one extending from nucleotide 199 to the end of the determined sequence. Within these latter approximately 150 nucleotides, there are four additional AUGs; two are in phase with the proposed reading frame, whereas two others occur in translation phases that are followed by termination codons.

The nucleotide sequence to which the leader

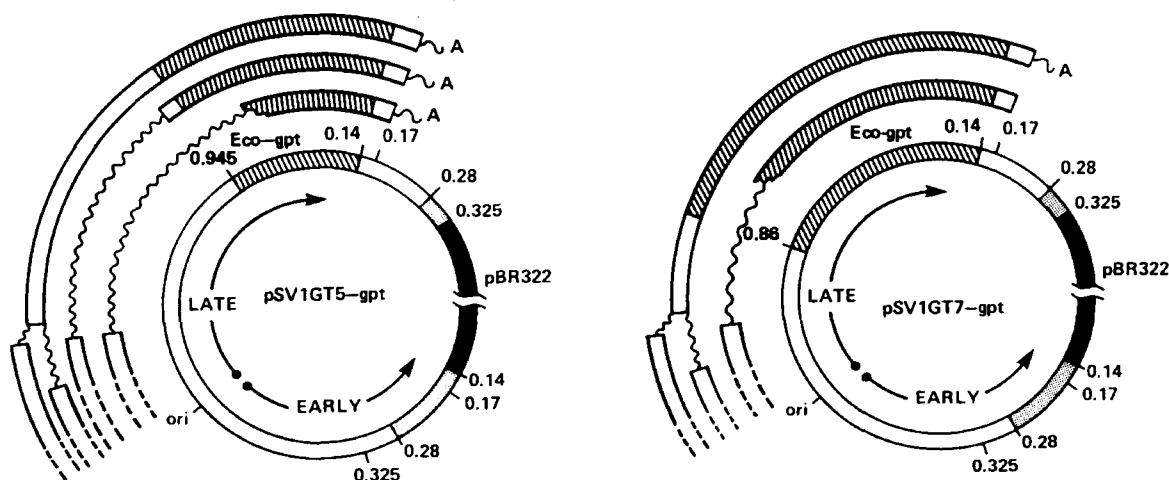


FIG. 5. Summary of the structures of pSV-*gpt* RNAs. The wavy lines indicate that the RNAs are composite structures composed of SV40 leader sequences linked to RNA segments containing the coding regions. The dashed endpoints of the leader sequences are intended to indicate the heterogeneity of the 5' ends (6). The truncated endpoint of the *gpt*-derived sequences in the small RNAs produced by pSV1GT5- and pSV1GT7-*gpt* represents the approximate location at which the SV40 leader has been spliced into the *gpt* segment.

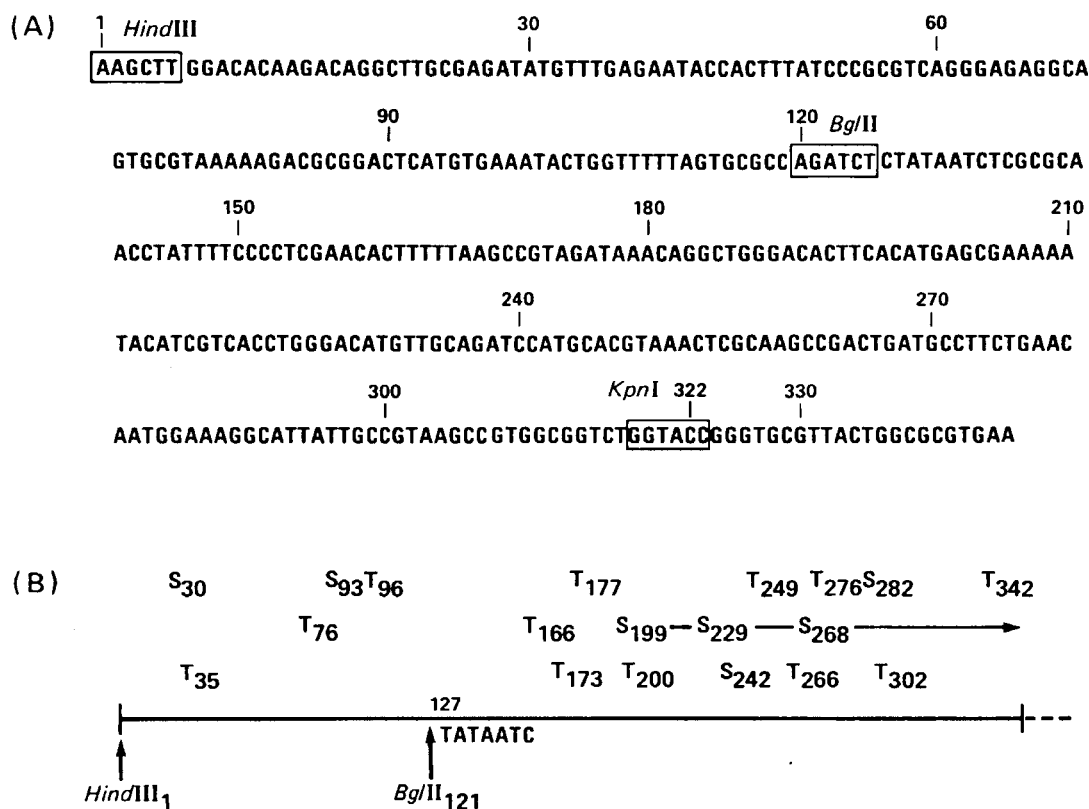


FIG. 6. 5'-Nucleotide sequences of *gpt*. (A) The *Hind*III, *Bgl*II, and *Kpn*I endonuclease cleavage sites are indicated. (B) The letters S and T represent AUG and any one of the known termination codons, respectively; the subscripts indicate the distance of the codon (in nucleotides) from the 5' end. The 7-nucleotide sequence adjacent to the *Bgl*II site locates a putative *E. coli* Pribnow box (25).

is spliced in the novel 1.2-kb RNA has not been determined, but the size of the *gpt* fragment protected in the S1 nuclease analysis indicates that the discontinuity occurs in the vicinity of the *Kpn*I site at nucleotide 322, more than 100 nucleotides beyond the putative start of the XGPRT coding sequence. Since the aberrant splice removes the start of the XGPRT coding sequence, we surmise that the novel mRNA does not produce active XGPRT.

Note that within the 198 nucleotides preceding the putative XGPRT coding sequence there are two AUG codons at nucleotides 30 and 93. However, each is followed by termination codons in all three translation phases. Thus, if the *gpt*-containing 16S-like mRNA produced in pSV1GT5-*gpt*-transfected cells contains the characteristic 200-nucleotide leader segment spliced to the SV40 sequence 5' proximal to the *gpt* segment (6), then the XGPRT initiation codon is more than 400 nucleotides from the 5' terminus of the RNA; moreover, that AUG codon is preceded by three AUGs, one in the leader segment and two in the *gpt* sequence. The corresponding 19S-like mRNA produced in

pSV1GT7-*gpt*-transfected cells is also translated to XGPRT; in this case the start of the *gpt* coding sequence may be as much as 700 to 850 nucleotides from the 5' terminus of the mRNA and be preceded by five AUGs, one in the leader, two in the VP2-VP3 coding sequence, and two in the *gpt* segment.

Deletions within the *gpt* sequence which affect the level of XGPRT production. To evaluate the assignment of the AUG at nucleotide 199 as the initiation codon for XGPRT and to evaluate the importance of the *gpt* 5'-untranslated region for expression of XGPRT, the first 121 or 322 nucleotides were deleted from the *gpt* segment in pSV2-*gpt* (see Fig. 7 and Materials and Methods). The variant plasmid DNAs were introduced into Gpt⁻ *E. coli* (GP120) to determine whether they still expressed *gpt* (21). GP120 can grow in minimal media supplemented with guanine but not in its absence; cells that acquire pSV2-*gpt* grow well in the absence of guanine (21). GP120 cells that acquire pSV2-*gpt* lacking the 5' segment up to the *Kpn*I restriction site [pSV2-*gpt* Δ(*Hind*III-*Kpn*I)] were Gpt⁻; those that received pSV2-*gpt* with the deletion

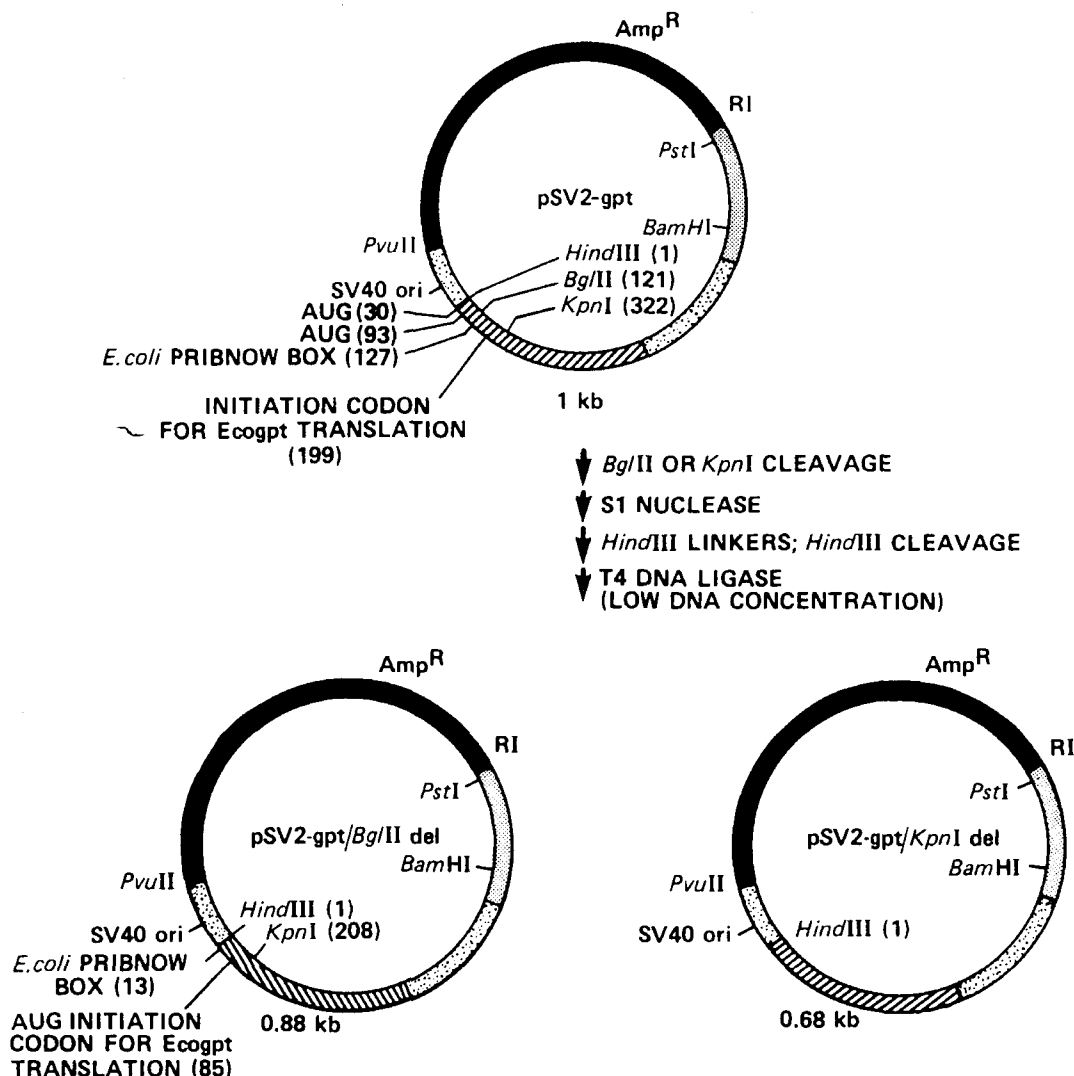


FIG. 7. Deletions of *gpt* sequences which affect XGPRT expression. See the text for the details of the construction of the deletions. The numbers in parentheses indicate the distance in nucleotides from the 5' end of the *gpt* fragment.

extending from the 5' end to the *Bgl*II cleavage site [pSV2-*gpt* Δ (*Hind*III-*Bgl*II)] grew slowly. These results suggest that in *E. coli* the first 121 nucleotides of the *gpt* DNA segment are needed for high-level expression of *gpt* and that the next 200 are essential for any expression. Quite possibly, removal of the first 121 residues in the sequence reduces transcription and hence *gpt* expression by deleting the -35 region upstream of the Pribnow sequence (27). Probably, the larger deletion inactivates *gpt* expression because it removes the promoter and amino-terminal coding region of *gpt*.

The consequence of these deletions on the expression of *gpt* in animal cells was determined by measuring the amount of *E. coli* XGPRT

made after transfection of CV1 cells with pSV2-*gpt*, pSV2-*gpt* Δ (*Hind*III-*Bgl*II), and pSV2-*gpt* Δ (*Hind*III-*Kpn*I) (Fig. 8). As was the case with the bacterial host, deletion of the first 322 nucleotides of the *gpt* DNA segment abolished XGPRT production. However, unlike the findings in bacteria, after transfection with pSV2-*gpt* Δ (*Hind*III-*Bgl*II) DNA, the production of *E. coli* XGPRT is approximately 5 to 10 times higher than that found with the unmodified pSV2-*gpt*.

DISCUSSION

Our experiments demonstrate that a bacterial gene, *gpt* (Ecogpt), can be expressed in mam-

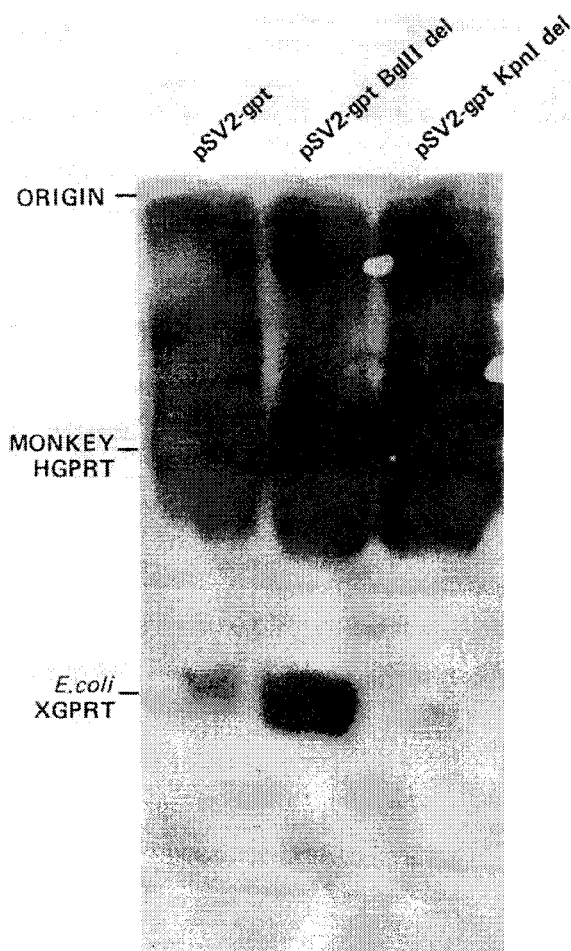


FIG. 8. Electrophoretic assay of GPRT in extracts of cells transfected with DNAs containing deletions in the *gpt* segment. Each track contained extract from about 10^6 transfected cells. *BglII* del, Δ (*HindIII*-*BglII*); *KpnI* del, Δ (*HindIII*-*KpnI*).

malian cells when it is incorporated into SV40 recombinant genomes that provide the signals needed for transcription and translation. Transcription of *gpt* in the pSV1 vector constructions probably occurs from the late promoters, since these mRNA's appear to contain leaders spliced to segments that contain the XGPRT coding sequence. pSV1GT5-*gpt*, which contains the inserted segment between SV40 map positions 0.945 and 0.145, retains the splice boundaries needed for the formation of 16S- and 19S-type mRNA's. As expected, corresponding sized spliced mRNA's containing the *gpt* segment are produced. With pSV1GT7-*gpt*, the *gpt* segment is inserted between SV40 map positions 0.86 and 0.145, and, therefore, a splicing sequence for 16S-type mRNA is eliminated, but the splicing sequence for 19S-type mRNA is retained. Not surprisingly, the 16S-type *gpt* mRNA is not

formed, but the corresponding 19S mRNA is readily detectable. Both recombinant genomes also give rise to a novel mRNA, one that seems to have a late leader spliced to a sequence in the *gpt* coding region. Sequence analysis of this species of *gpt*-containing mRNA is needed to identify the leader segment and the splice junctions in this novel mRNA.

Although the splicing noted here is the first recorded example of splicing from a eucaryote to a procaryote RNA sequence, aberrant splicing has also been detected with other recombinants. SVGT7- β -globin recombinants also yield two spliced mRNA's; one corresponds to the expected 19S-type mRNA, and the other contains a leader segment spliced to the 5'-untranslated segment of β -globin cDNA (23; J. L. Sklar and P. Berg, unpublished data). Moreover, A. R. Buchman and P. Berg (unpublished data) and Chu and Sharp (4a) have shown that splicing can occur between unrelated donor and acceptor splice junctions. Additionally, SV40 leader segments are also spliced to the acceptor junction of the β -globin large intervening sequence (D. Canaani and P. Berg, unpublished data) and to the rat insulin I intervening sequence (S. C. Clark, C. Nguyen-Huu, H. M. Goodman, and P. Berg, unpublished data). What nucleotide sequences signal splicing and how these are affected by neighboring sequences have yet to be explained.

A major consideration in our earlier design of SV40 recombinants was to locate protein coding sequences within an RNA so that they would be expressed. In constructing SV40- β -globin recombinants (22), the β -globin cDNA segment was introduced into the SV40 genome at a position nearly identical to that occupied by the VP1 coding sequence; our expectation was that this arrangement would promote the most efficient translation of the resulting SV40- β -globin mRNA's. But because the *gpt* DNA segment was isolated solely by its ability to complement the growth of *Gpt*⁻ *E. coli*, the actual location of the codon for initiating translation of XGPRT and the nature of the preceding sequences were unknown. The nucleotide sequence of the 5' region of the *gpt* segment indicates that the most likely initiation codon for XGPRT translation is located 200 nucleotides from the 5' end of the 1-kb cloned fragment and that it is preceded by two AUG codons. Nevertheless, mRNA's produced from each of the pSV-*gpt* recombinants are translated to yield what appears to be the bacterial enzyme. In the most extreme case, the putative XGPRT initiator AUG in pSV1GT7-*gpt* RNA is located 700 to 850 nucleotides from the 5' terminus of the 19S-type leader segment

and is preceded by five AUG codons: one AUG in the leader segment, one each corresponding to the beginning of the VP2 and VP3 coding sequences, and two in the *gpt* segment. These results indicate that a procaryote translation initiation signal can be recognized in animal cells; there also appears to be no block to the initiation of translation at the internal XGPRT initiator AUG. Although the translation of XGPRT from minor, aberrantly spliced RNAs cannot be rigorously excluded, we believe that this is unlikely. The same paradoxical translation of XGPRT occurs with mRNA produced from pSV2-*gpt*, which contains an intact *gpt* DNA segment and, therefore, two AUG codons before the XGPRT initiator AUG. Additional examples of translation initiated on internal AUG codons have also been noted with other recombinants. For example, after transfection of monkey cells with pSV1GT7-*dhfr* cDNA, dihydrofolate reductase is produced at high levels from only 19S-type spliced mRNA's, yet these contain the two AUG codons in the SV40 sequence 5' to the dihydrofolate reductase coding sequence (Subramani, Mulligan, and Berg, unpublished data). Nevertheless, when the β -globin cDNA sequence occurs within the VPI (23) or large T-antigen coding regions (P. J. Southern and P. Berg, unpublished data), β -globin is not synthesized from the mRNA's containing the β -globin sequence.

These findings bear on current notions regarding the general mechanism of initiation of eucaryote translation. The prevailing view is that with few exceptions, only the first AUG in eucaryote mRNA's serves as an initiator codon for translation (for review, see reference 14). Kozak has suggested (14, 15) a scanning mechanism to explain the findings with eucaryote mRNA. This involves attachment of the small ribosomal subunit at or near the 5' end of mRNA's and movement of the subunit in the 3' direction relative to the RNA sequence until the first AUG is reached. Then, at some inherent efficiency, the complete ribosome is assembled, and translation is initiated at that site. In natural eucaryote mRNA's, perhaps the first AUG functions efficiently to promote initiation, and significant internal initiation is pre-empted. However, it is possible that there are neighboring structural features in eucaryote mRNA's which favor initiation at the first AUG codons and are ordinarily absent or unavailable for initiation at internal AUG codons. For example, the relative ability of an internal AUG to interact with the 5'-capped terminus of the mRNA may be important in this regard (C. Queen, B. Paterson, and M. Rosenberg, private communication). The in-

creased efficiency of XGPRT production when the two upstream AUGs are removed from *gpt* are noteworthy with respect to both of these notions. The ability to construct novel arrangements of coding sequences in eucaryote mRNA's and to modify, selectively, portions of these nucleotide sequences provides a new way to explore the rules governing eucaryote mRNA translation.

The present studies suggest that the requirements for obtaining expression of a wide variety of procaryote genes are not formidable. Beside *gpt*, the phosphotransferase gene encoding neomycin resistance (Neo^r) from the bacterial transposon Tn5 (13) has been cloned in pSV2 and related vector DNAs and then expressed in a variety of cultural mammalian cells (Southern and Berg, unpublished data). There is no reason to believe that the bacterial genes, *gpt* and that coding for Neo^r, are unique in their ability to be expressed in mammalian cells. It seems more likely that bacteria, their viruses, and simple eucaryotes will provide a rich source of genes for the modification of mammalian cells.

ACKNOWLEDGMENTS

We thank Brian Nichols and Charles Yanofsky for sharing their expertise in nucleotide sequencing with us.

This research was performed with funds provided by the National Institute of General Medical Sciences (Public Health Service grant GM-13235), the National Cancer Institute (Public Health Service grant CA-15513-07), and the American Cancer Society (MV-35J). R.C.M. was a National Science Foundation predoctoral fellow.

LITERATURE CITED

1. Alwine, J. C., D. J. Kemp, and G. R. Stark. 1977. Methods for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U.S.A.* 74:5350-5354.
2. Berk, A. J., and P. Sharp. 1978. Spliced early mRNAs of simian virus 40. *Proc. Natl. Acad. Sci. U.S.A.* 75:1274-1278.
3. Chang, A. C. Y., J. H. Nunberg, R. J. Kaufman, H. A. Erlich, R. T. Schimke, and S. N. Cohen. 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature (London)* 275:617-624.
4. Chasin, L. A., and G. Urlaub. 1976. Mutant alleles for hypoxanthine phosphoribosyltransferase: codominant expression, complementation, and segregation in hybrid chinese hamster cells. *Somat. Cell Genet.* 2:453-467.
- 4a. Chu, G., and P. A. Sharp. 1981. A gene chimaera of SV40 and mouse β -globin is transcribed and properly spliced. *Nature (London)* 289:378-382.
5. Davis, R. W., M. Thomas, J. Cameron, T. P. St. John, S. Scherer, and R. Padgett. 1980. Rapid DNA isolations for enzymatic and hybridization analysis. *Methods Enzymol.* 65:404-411.
6. Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. V. Choudary, P. Leibowitz, and S. M. Weissman. 1978. The

- 5'-terminal leader sequence of late 16s RNA from cells infected with simian virus 40. *J. Biol. Chem.* **253**:3643-3647.
7. Goff, S. P., and P. Berg. 1976. Construction of hybrid viruses containing SV40 and λ phage DNA segments and their propagation in cultured monkey cells. *Cell* **9**: 695-705.
8. Goff, S. P., and P. Berg. 1979. Construction and expression of simian virus 40 recombinant genomes containing the *Escherichia coli* gene for thymidine kinase and a *Saccharomyces cerevisiae* gene for tyrosine transfer RNA. *J. Mol. Biol.* **133**:359-383.
9. Graham, R., and A. Van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52**:456-467.
10. Gruss, P., C.-J. Lai, R. Dhar, and G. Khoury. 1979. Splicing as a requirement for biogenesis of functional 16S mRNA of SV40. *Proc. Natl. Acad. Sci. U.S.A.* **76**: 4317-21.
11. Heyneker, H. L., J. Shine, H. M. Goodman, H. Boyer, J. Rosenberg, R. E. Dickerson, S. A. Narang, K. Itakura, S. Y. Lin, and A. D. Riggs. 1976. Synthetic lac operator DNA is functional *in vivo*. *Nature (London)* **263**:748-752.
12. Holden, J. A., P. D. Harriman, and J. D. Wall. 1976. *Escherichia coli* mutants deficient in guanine-xanthine phosphoribosyl transferase. *J. Bacteriol.* **126**:1141-1148.
13. Jorgensen, R. A., S. J. Rothstein, and W. S. Reznikoff. 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Mol. Gen. Genet.* **177**:65-72.
14. Kozak, M. 1978. How do eukaryote ribosomes select initiation regions in messenger RNA? *Cell* **15**:1109-1123.
15. Kozak, M. 1980. Evaluation of the scanning model for initiation of protein synthesis in eukaryotes. *Cell* **22**:7-8.
16. Krenitsky, T. A., R. Papaioanou, and G. B. Elion. 1969. Human hypoxanthine phosphoribosyltransferase: purification, properties, and specificity. *J. Biol. Chem.* **244**:1263-1270.
17. Mandel, M., and A. Higa. 1970. Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**:159-162.
18. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**:560-564.
19. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. U.S.A.* **74**:4835-4838.
20. Miller, R. L., G. A. Ramsay, T. A. Krenitsky, and G. Elion. 1972. Guanine phosphoribosyltransferase from *Escherichia coli*, specificity and properties. *Biochemistry* **11**:4723-4731.
21. Mulligan, R. C., and P. Berg. 1980. Expression of a bacterial gene in mammalian cells. *Science* **209**:1422-1427.
22. Mulligan, R. C., B. H. Howard, and P. Berg. 1979. Synthesis of rabbit β -globin in cultured monkey kidney cells following infection with a SV40 β -globin recombinant genome. *Nature (London)* **277**:108-114.
23. Mulligan, R. C., R. T. White, and P. Berg. 1980. Formation of β -globin following infection with recombinants containing rabbit β -globin cDNA at different locations of SV40's late region, p. 201-215. *In* 17th Miami Winter Symposium: mobilization and reassembly or genetic information. Academic Press, Inc., New York.
24. Parker, B. A., and G. R. Stark. 1979. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections by virus or virus DNA. *J. Virol.* **31**:360-369.
25. Pribnow, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. U.S.A.* **72**:784-788.
26. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.
27. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the initiation and termination of RNA transcription. *Annu. Rev. Genet.* **13**:319-353.
28. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure. *Nucleic Acids Res.* **75**:1175-1193.
29. Weinberg, R., Z. Ben-Ishai, and J. Newbold. 1974. Simian virus 40 transcription in productively infected and transformed cells. *J. Virol.* **13**:1263-1273.